

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/65747 A1

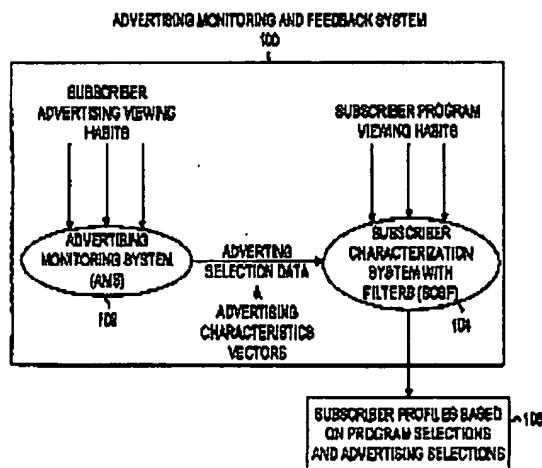
- (51) International Patent Classification: H04H 9/00, H04N 7/025, 7/10, 7/16 (74) Agents: RYDER, Douglas, J. et al.; 300 North Broad Street, Doylestown, PA 18901 (US).
- (21) International Application Number: PCT/US01/06459 (81) Designated States (national): AB, AO, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KB, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (22) International Filing Date: 1 March 2001 (01.03.2001) (84) Designated States (regional): ARIPO patent (GH, GM, KB, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TO).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/516,983 1 March 2000 (01.03.2000) US
09/516,914 1 March 2000 (01.03.2000) US
- (71) Applicant (for all designated States except US): EXPANSE NETWORKS, INC. (US/US); 300 North Broad Street, Doylestown, PA 18901 (US).
- (72) Inventor; and
- (73) Inventor/Applicants (for US only): ELDERING, Charles, A. (US/US); 214 Commons Way, Doylestown, PA 18901 (US). SYLLA, M., Lamine (SN/US); 347 East Butler Avenue, Apt. C, New Britain, PA 18901 (US). BLASKO, John, P. (US/US); 4 Old Mill Lane, New Hope, PA 18938 (US).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

[Continued on next page]

(54) Title: ADVERTISEMENT MONITORING AND FEEDBACK SYSTEM



(57) Abstract: An advertising monitoring and feedback system (100) is presented in which subscriber selections including channel changes are monitored, and in which information regarding an advertisement is extracted from text related to the advertisement. The text related to the advertisement is in the form of closed-caption text, data transmitted with the advertisement, or other associated text. A record of the effectiveness of the advertisement is created in which measurements of the percentage of the advertisement which was viewed are stored. Such records are then fed back to a subscriber characterization system (104), thereby the subscriber characterization system (104) generates subscriber profiles (106) that incorporate advertising related information. Prior to generating subscriber profiles (106), the subscriber characterization system (104) also evaluates the subscriber selection data to eliminate data associated with irrelevant activities, such as channel surfing, channel jumping, or extended periods of inactivity.

TITLE***Advertisement Monitoring And Feedback System*****Background of the Invention**

5 Cable television service providers have typically provided one-way broadcast services but now offer high-speed data services and can combine traditional analog broadcasts with digital broadcasts and access to Internet web sites. Telephone companies can offer digital data and video programming on a
10 switched basis over digital subscriber line technology. Although the subscriber may only be presented with one channel at a time, channel change requests are instantaneously transmitted to centralized switching equipment and the subscriber can access the programming in a broadcast-like
15 manner. Internet Service Providers (ISPs) offer Internet access and can offer access to text, audio, and video programming which can also be delivered in a broadcast-like manner in which the subscriber selects "channels" containing programming of interest. Such channels may be offered as part
20 of a video programming service or within a data service and can be presented within an Internet browser.

 Advertisements are a part of daily life and certainly an important part of entertainment programming, where the payments for advertisements cover the cost of network television. A
25 method, which provides a flexible billing plan to cable network users based on the amount of advertisements viewed is described in U.S. Patent No. 5,532,735, which discloses a method of advertisement selection for interactive services. A user associated with an interactive TV is presented with a program
30 and a set of advertisements. The user can indicate the amount of advertisements in the set of advertisements he wants to view.

 While advertisements are sometimes beneficial to subscribers and deliver desired information regarding specific

U.S. Patent No. 5,608,445 discloses a method and device for data capture in television viewer research. Devices are attached to a video installation in order to determine to which channel a set is tuned.

5 With the advent of the Internet, manufacturers and service providers have found ways to selectively insert their advertisements based on a subscribers requests for information. As an example, an individual who searches for "cars" on the Internet may see an advertisement for a particular type of car.

10 Various internet-based advertisers use this method. The product literature from IMGIS Inc., "Ad Force," printed from the World Wide Web site http://www.starpt.com/core/ad_Target.html on June 30, 1998 discloses an ad targeting system. The system delivers ads to

15 web site visitors based on the content of the web page, time of day, day of the week, keyword, by the number of times a visitor sees an advertisement and by the order in which a series of advertisements are shown to a visitor. Nevertheless, unless the subscriber actually goes to the advertised web site, there

20 is no way to determine if the advertisement has been watched. As the content on the Internet migrates to multimedia programming including audio and video, the costs for the advertising will increase, but unless the advertiser can be sure that a significant percentage of the message was watched

25 or observed, the advertising is ineffective. Prior art products for generating reports of ad campaigns are generally PC-centric as described in various product literature which include the product literature from DoubleClick Inc., "DoubleClick: Reporting," printed from the World Wide Web (WWW)

30 site http://www.doubleclick.net/dart/howl_repo.htm on June 19, 1998, which discloses the reporting capabilities of DoubleClick's Dynamic Advertising Reporting & Targeting (DART) product. The information in the reports includes daily impressions by advertisement type, average impression per day

In order to deliver more targeted programming and advertising to subscribers, it is necessary to understand their likes and dislikes to a greater extent than is presently done today. Systems which identify subscriber preferences based on their purchases and responses to questionnaires allow for the targeted marketing of literature in the mail, but do not in any sense allow for the rapid and precise delivery of programming and advertising which is known to have a high probability of acceptance to the subscriber. Other systems give users the possibility to choose their programming as described in U.S. Patent No. 5,223,924 which discloses a system and method for automatically correlating user preferences with a TV program information database. The system includes a processor that performs "free text" search techniques to correlate the downloaded TV program information with the viewer's preferences. This system requires an interaction between the users and the programming. The white paper from Net Perceptions corporation entitled "Adding Value in the Digital Age" and printed from the World Wide Web site <http://www.netperceptions.com/products/white-papers.html> on June 30, 1998 discloses how the GroupLens Recommendation Engine gives online businesses the ability to target and personalize services, content, products and advertising. A learning process learns personal information about an individual using explicit and implicit ratings, a prediction process predicts user preference using collaborative filtering and the recommendation process recommends products or services to users based on predictions.

The product literature from Aptex Software Inc., "SelectCast for Commerce Servers," printed from the World Wide Web site <http://www.aptex.com/products-selectcast-commerce.htm> on June 30, 1998 describes the product "SelectCast" for Commerce Servers. It personalizes online shopping based on observed user behavior. User interests are learned based on

relationship-based context vectors based on word proximity and co-importance. Geometric relationships among context vectors are representative of conceptual relationships among their associated items.

- 5 The product data sheet from Open Sesame, "Learn Sesame," printed from the World Wide Web site http://www.opensesame.com/prod_04.html on July 09, 1998 discloses Open Sesame's personalization product for Web enterprises. It learns about users automatically from their
- 10 browsing behavior.

- The product literature from Engage Technologies, "Engage.Discover," printed from the World Wide Web site <http://www.engagetechnologies.com> on July 09, 1998 discloses Engage Technologies' product for user profiling. User-disclosed
- 15 information such as interest, demographics and opinions are combined with anonymous clickstream data that describes where users come from before visiting the site, how long they stay, and what pages or types of pages they visit most frequently to build the visitor profile.

- 20 The marketing literature from BroadVision, "The Power of Personalization", printed from the World Wide Web site <http://www.broadvision.com/content/corporate/brochure/Broch4.htm> on August 21, 1998 discloses the BroadVision One-to-One application profiling system. The system learns about users
- 25 through a variety of techniques including registration, questionnaires, observation and integration of historical and externally generated data.

- The marketing literature from the Firefly Corporation, "Firefly Passport Office," printed from the World Wide Web site
- 30 <http://www.firefly.net/company/PassportOffice.html> on June 20, 1998 discloses Firefly's Relationship Management software. The software enables online businesses to create, extend and manage personal profiles for every user.

system. The EPG data includes guide data, channel data and program data. The program data includes among other information, the program title, the program category, the program sub-category and a detailed description of the program.

5 Some efforts have been made to transmit targeted advertisements and thereupon collect feedback on the transmitted advertisements. For example, U.S. Patent No. 5,948,061 provides methods and apparatuses for targeting the delivery of advertisements over a network such as the Internet.

10 Statistics are compiled on individual users and networks and the use of the advertisements is tracked to permit targeting of the advertisements to individual users. In response to requests from affiliated sites, an advertising server transmits to people accessing the page of a site an appropriate one of

15 the advertisements based upon profiling of users and networks.

 Furthermore, U.S. Patent No. 6,005,597 provides a method and apparatus for television program selection that monitors the viewing preferences of a viewer to create a dynamic viewer profile that is used to rate available programs. Based on the

20 viewer profile, available programs are sorted and presented to the viewer in descending order of predicted interest. The invention allows a viewer to quickly find the program of greatest interest to the viewer without having to tediously search through large numbers of available programs. The

25 invention may also be used for selecting from among a plurality of programs other than television programs, such as, for example, radio programs or audio or video programs stored on digital storage media such as CD's and DVD's.

 U.S. Patent No. 5,446,919 provides a communication system

30 capable of targeting a demographically or psychographically defined audience. A master database is maintained, containing demographic and psychographic information about each audience member. This information is transmitted and stored in a channel selection/decoder unit associated with each audience

a remote control, downloaded to a subscriber's converter from a remote head-end, or programmed into the converter at installation. Prioritization of the demographic characteristics of a plurality of television viewers watching a program together enables commercials to be targeted to the viewer having highest priority. Statistical data can be maintained concerning the number and identity of subscribers viewing specific commercials.

U.S. Patent No. 5,774,170 enhances television (and radio) advertising by targeting, delivering and displaying electronic advertising messages (commercials) within specified programming in one or more pre-determined households (or on specific display devices) while simultaneously preventing a commercial from being displayed in other households or on other displays for which it is not intended. Commercials can be delivered to specified homes or displays via either over-the-air or wired delivery systems.

U.S. Patent No. 5,446,919 provides a communication system capable of targeting a demographically or psychographically defined audience. A master database is maintained, containing demographic and psychographic information about each audience member. This information is transmitted and stored in a channel selection/decoder unit associated with each audience member's receiver. Multiple media messages are transmitted to each audience member. Accompanying the transmission is a selection profile command, which details the demographic/psychographic profile of audience members that are to receive each media message. The channel selector/decoder unit associated with each member's receiver compares the selection profile with the demographic/psychographic information about the audience member and selects the appropriate media message for that audience member.

IEEE Publication: "A Framework for Targeting Banner Advertising on the Internet" by K.Gallagher and J. Parsons,

and generates one or more records of the subscriber advertising selection data. In a preferred embodiment, the subscriber advertising selection data includes the channel selected and the time at which the channel was selected. Advertisement
5 related information including the type of product, brand name, and other descriptive information that categorizes the advertisement is also extracted and added to the advertising selection data. The advertisement related information might also be extracted from the closed-captioning text. Based on
10 this advertisement selection data, a record is created documenting what percentage, if any, of the advertisement was watched. This record is then fed back to the subscriber characterization system, wherein the subscriber characterization system may comprise one or more filters.

15 The subscriber characterization system also monitors the subscriber viewing habits associated with program viewing and generates one or more records of program selection choices. The program selection choices include the viewing time duration, number of channel changes, volume at which the
20 programming is listened, program selection, and text information about the programming. The records are used to determine what type of programming the subscriber is most interested in. The subscriber characterization system characterizes subscribers based on feedback information from an
25 advertisement monitoring system and monitors their detailed program viewing selections.

Furthermore, the subscriber characterization system is equipped with one or more filters that assist in determining and eliminating from consideration, selection data associated
30 with irrelevant activities by the subscriber. Examples of irrelevant activities include selection data associated with channel surfing and/or channel jumping (up and down) activities.

The channel surfing activity refers to one or more rapid

programming. The reasons for such dead periods may be caused by the fact that the subscriber has left the room, or the subscriber is not active (e.g., the subscriber has gone to sleep or has dozed off), or the fact that the subscriber is
5 actively engaging in another activity within the room and is not attending to the programming.

The subscriber characterization system of the present invention analyzes the actual viewing selections made by the subscriber or the subscriber household, and generates a
10 demographic description of the subscriber or household. This demographic description describes the probable age, income, gender and other demographics. The resulting characterization includes probabilistic determinations of what other programming or products in which the subscriber/household will be
15 interested.

In a preferred embodiment, the text information related to the advertisement is processed using context mining techniques which allow for classification of the advertisement and extraction of key data including product type and brand.
20 Context mining techniques allow for determination of a product type, product brand name and in the case of a product which is not sold with a particular brand name, a generic name for the product.

The present invention can also be realized in a client-server mode in which case the advertising related information is collected at the client side of the network and is then transmitted to the server side via a secure connection. The server side then incorporates this information with other subscriber information to create subscriber profiles.
25

30 These and other features and objects of the invention will be more fully understood from the following detailed description of the preferred embodiments which should be read in light of the accompanying drawings.

FIG. 8B illustrates a flowchart for program characterization;

FIGS. 9A illustrates a deterministic program category vector;

5 FIG. 9B illustrates a deterministic program sub-category vector;

FIG. 9C illustrates a deterministic program rating vector;

FIG. 9D illustrates a probabilistic program category vector;

10 FIG. 9E illustrates a probabilistic program sub-category vector;

FIG. 9F illustrates a probabilistic program content vector;

FIG. 10A illustrates a set of logical heuristic rules;

15 FIG. 10B illustrates a set of heuristic rules expressed in terms of conditional probabilities;

FIG. 11 illustrates an entity-relationship diagram for the generation of program demographic vectors;

FIG. 12 illustrates a program demographic vector;

20 FIG. 13 illustrates an entity-relationship diagram for the generation of household session demographic data and household session interest profiles;

FIG. 14 illustrates an entity-relationship diagram for the generation of average and session household demographic characteristics;

FIG. 15 illustrates average and session household demographic data;

FIG. 16 illustrates an entity-relationship diagram for generation of a household interest profile;

30 FIG. 17 illustrates a household interest profile including programming and product profiles;

FIG. 18 illustrates a client-server architecture for realizing the present invention; and

FIG. 19 illustrates an advertisement monitoring table.

as they relate to actual programming information. SCSF 104, based on advertising related information and the actual programming information, generates one or more subscriber profiles 106, wherein each subscriber profile indicates a probabilistic measure of subscriber demographics and/or preferences.

As illustrated in FIG. 1B, in generating one or more records of advertisement selections, the AMS 102 receives from a subscriber 120A commands in the form of a volume control signal 124A or advertising selection data 122A which can be in the form of a channel change. The advertising material being viewed by the subscriber 120A is referred to as source material 130A. The source material 130A, as defined herein, is the content that a subscriber selects and may consist of analog video, Motion Picture Expert Group (MPEG) digital video source material, other digital or analog material, Hypertext Markup Language (HTML) or other type of multimedia source material. The AMS 102 can access the source material 130A received by the subscriber 120A using a start signal 132A and a stop signal 134A, which control the transfer of source related text 136A which can be analyzed as described herein.

In a preferred embodiment, the source related text 136A can be extracted from the source material 130A and stored in memory. The source related text 136A, as defined herein, includes source related textual information including descriptive fields which are related to the source material 130A, or text which is part of the source material 130A itself. The source related text 136A can be derived from a number of sources including but not limited to closed-captioning information, Electronic Program Guide (EPG) material, and text information in the source itself (e.g. text in HTML files).

An Electronic Program Guide (EPG) 140A contains information related to the source material 130A which is useful to the subscriber 120A. The EPG 140A is typically a

In the process of collecting raw subscriber selection data, the SCSF 104 receives from a subscriber 120B commands in the form of a volume control signal 124B or program selection data 122B which can be in the form of a channel change but may
5 also be an address request which requests the delivery of programming from a network address. A record signal 126B indicates that the programming or the address of the programming is being recorded by the subscriber 120B. The record signal 126B can also be a printing command, a tape
10 recording command, a bookmark command or any other command intended to store the program being viewed, or program address, for later use.

The material being viewed by the subscriber 120B is referred to as source material 130B. The source material 130B,
15 as defined herein, is the content that a subscriber selects and may consist of analog video, digital video such as MPEG digital video, Hypertext Markup Language (HTML), or other types of multimedia source material. The SCSF 104 can access the source material 130B received by the subscriber 120B using a start
20 signal 132B and a stop signal 134B, which control the transfer of source related text 136B which can be analyzed as described herein.

In a preferred embodiment, the source related text 136B can be extracted from the source material 130B and stored in
25 memory. The source related text 136B, as defined herein, includes source related textual information including descriptive fields which are related to the source material 130B, or text which is part of the source material 130B itself. The source related text 136B can be derived from a number of
30 sources including but not limited to closed-captioning information, EPG material, and text information in the source itself (e.g., text in HTML files).

The EPG 140B contains information related to the source material 130B which is useful to the subscriber 120B. The EPG

selection data associated with channel surfing, channel jumping and dead periods is eliminated from the advertisement and program selection data to generate actual subscriber selection data 199. FIG. 1D and the related text describe this process
5 in detail.

The actual subscriber selection data 199 comprises time 112C, which corresponds to the time of an actual viewing event exclusive of channel surfing, channel jumping or dead periods, channel ID 114C, program ID 116C, program title 117C, volume
10 level 118C, and channel change record 119C.

Based on the actual subscriber selection data 199, the SCSF 104 generates one or more program characteristics vectors 150 which are comprised of collected characteristics data 152. The characteristics data 152, which can be used to create the
15 program characteristics vectors 150, both in vector and table form, are examples of source related information which represent characteristics of the source material (for the advertisements and the programs). The characteristics vectors 150 are derived from the source related text 136B and/or from
20 the EPG 140B by applying information retrieval techniques. The details of this process are discussed in detail later.

In a preferred embodiment, the characteristics vectors 150 are lists of values which characterize the programming (source) material in accordance to the category 144B, the sub-category
25 146B, and the program description 148B. The present invention may also be applied to advertisements, in which case program characteristics vectors contain, as an example, a product category, a product sub-category, and a brand name.

As illustrated in FIG. 1C, the SCSF 104 uses heuristic
30 rules 160. The heuristic rules 160, as described herein, are composed of both logical heuristic rules as well as heuristic rules expressed in terms of conditional probabilities. The heuristic rules 160 can be accessed by the SCSF 104 via a request rules signal 162 which results in the transfer of a

generate actual subscriber selection data 199 which does not include irrelevant selection data. The irrelevant selection data generally corresponds to channel surfing, channel jumping, or dead periods activities. These activities are generally
5 recognized by reviewing corresponding viewing times. In the case of channel surfing or channel jumping, the associated viewing times are very brief, a few milliseconds or a few seconds. In the case of dead periods, the viewing time is relatively long having no actions, e.g., a few hours.

10 FIG. 2 illustrates an exemplary system for monitoring subscriber activities, including advertising viewing habits, and can be used to realize the AMS 102 and the SCSF 104. In a preferred embodiment, the monitoring system of FIG. 2 is located in a television set-top device or in the television
15 itself. In an alternate embodiment, the monitoring system is part of a computer which receives programming from a network.

In an application of the system for television services, an input connector 220 accepts the video signal coming either from an antenna, cable television input, or other network. The
20 video signal can be analog or digital, such as MPEG. Alternatively, the video source may be a video stream or other multimedia stream from a communications network including the Internet.

In the case of either analog or digital video, selected
25 fields are defined to carry EPG data or closed-captioning text. For analog video, the closed-captioning text is embedded in the vertical blanking interval (VBI). As described in US Patent 5,579,005, assigned to Scientific-Atlanta, Inc., the EPG information can be carried in a dedicated channel or embedded
30 in the VBI. For digital video, the closed-captioning text is carried as video subscriber bits in a subscriber_data field. The EPG data is transmitted as ancillary data and is multiplexed at the transport layer with the audio and video data.

call sign, program name and other useful data are provided. In a preferred embodiment, the data stream is stored in a reserved location of the RAM 204. In an alternate embodiment, a magnetic disk is used for data storage. The system control unit 200 writes in a dedicated memory, which in a preferred embodiment is the RAM 204, the selected channel, the time of selection, the volume level and the program ID and the program title. Upon receiving the program selection data, the new selected channel is directed to the channel processor 210 and the system control unit 200 writes to the dedicated memory the channel selection end time and the program title at the time of channel change. The system control unit 200 keeps track of the number of channel changes occurring during the viewing time via the channel change record. This data forms part of the advertising selection records 110A or the programming selection records 110B.

The volume control signal is sent to the audio processor 240. In a preferred embodiment, the volume level selected by the subscriber corresponds to the listening volume. In an alternate embodiment, the volume level selected by the subscriber represents a volume level to another piece of equipment such as an audio system (home theatre system) or to the television itself. In such a case, the volume can be measured directly by a microphone or other audio sensing device which can monitor the volume at which the selected source material is being listened.

A program change occurring while watching a selected channel is also logged by the system control unit 200. Monitoring the content of the program at the time of the program change can be done by reading the content of the EDS. The EDS contains information such as program title, which is transmitted via the VBI. A change on the program title field is detected by the monitoring system and logged as an event. In an alternate embodiment, an EPG is present and program

FIG. 4 illustrates an exemplary block diagram of a computer system for a realization of the subscriber monitoring system illustrated in FIG.2. A system bus 422 transports data amongst the CPU 203, the RAM 204, a Read Only Memory - Basic Input Output System (ROM-BIOS) 406 and other components. The CPU 203 accesses a hard drive 400 through a disk controller 402. The standard input/output devices are connected to the system bus 422 through the I/O controller 201. A keyboard is attached to the I/O controller 201 through a keyboard port 416 and the monitor is connected through a monitor port 418. The serial port device uses a serial port 420 to communicate with the I/O controller 201. Industry Standard Architecture (ISA) expansion slots 408 and Peripheral Component Interconnect (PCI) expansion slots 410 allow additional cards to be placed into the computer. In a preferred embodiment, a network card is available to interface a local area, wide area, or other network.

FIG. 5 illustrates a channel sequence and volume over a twenty-four (24) hour period associated with advertisement selection record 110A or subscriber selection record 110B. The Y-axis represents the status of the receiver in terms of on/off status and volume level. The X-axis represents the time of day. The channels viewed are represented by the windows 501-506, with a first channel 502 being watched followed by the viewing of a second channel 504, and a third channel 506 in the morning. In the evening a fourth channel 501 is watched, a fifth channel 503, and a sixth channel 505. A channel change is illustrated by a momentary transition to the "off" status and a volume change is represented by a change of level on the Y-axis.

A detailed record of the advertising selection record 110A and/or programming selection data 110B is illustrated in FIG. 6A in a table format. A time column 602 contains the starting time of every event occurring during the viewing time. A

Similarly, the viewing record also indicates that the corresponding viewing times of each of channel numbers 7, 8, 9, 58, 57, 56, 55, 54, 53 are about a minute or less. This implies that after the subscriber had completed the viewing of channel number 6, the subscriber once again surfed the channels to find a programming of interest at channel 25, which was viewed for about 10 minutes.

FIG. 6C illustrates processing involved in the elimination of viewing times associated with the channel jumping activities. The channel jumping activity is different than a channel surfing activity in a sense that the subscriber already knows the intended programming (and corresponding channel number) he wants to watch, and utilizes the channel up or channel down button to arrive at the intended channel.

The viewing time of all the intermediate channels during channel jumping activity are generally very brief (less than a second). Also, as the channel up or channel down button is utilized to reach the desired channels, generally, there exists an upwards or a downwards stream of channel changes, i.e., the subscriber may jump through channels 2, 3, 4 and 5 to reach channel number 6 (an intended channel). Similarly, subscriber jumps may through channel 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16 to reach channel 17.

The filters of the present invention are configured to eliminate the channel jumping data from the actual viewing data. Filters generally evaluate the associated viewing times, and all the viewing times which correspond to channel jumping, e.g., are less than one second, are removed from the viewing records.

The filters are also configured to eliminate data associated with dead activities, e.g., extended spans of inactivity. These extended spans of inactivity indicate that the subscriber is not actively watching the programming, e.g., the subscriber has left the room, has gone to sleep, or is

statistical record contains the totals for the items listed in the minutes watched column 702, the channel changes column 704 and the average volume 706.

FIG. 8A illustrates an entity-relationship diagram for the generation of the characteristics vector 150. The context vector generation and retrieval technique can be applied for the generation of the characteristics vectors 150 which is described in US Patent 5,619,709. Other techniques are well known by those skilled in the art.

The source material 130 or the EPG 140 is passed through a characterization processing 800 to generate the characteristics vectors 150. The characterization processing 800 is described in accordance with FIG. 8B. The content descriptors including a first content descriptor 802, a second content descriptor 804 and an nth content descriptor 806, each classified in terms of the category, the sub-category, and other divisions as identified in the industry accepted program classification system, are presented to a context vector generator 820. As an example, the content descriptor can be text representative of the expected content of material found in the particular category (program and/or advertisement). In this example, the content descriptors 802, 804 and 806 would contain text representative of what would be found in programs in the news, fiction, and advertising categories respectively. The context vector generator 820 generates context vectors for that set of sample texts resulting in a first summary context vector 808, a second summary context vector 810, and an nth summary context vector 812. In the example given, the summary context vectors 808, 810, and 812 correspond to the categories of news, fiction and advertising respectively. The summary vectors are stored in a local data storage system.

A sample of the source related text 136 which is associated with the new program to be classified is passed to the context vector generator 820 which generates a context

The program characteristics vectors 150 in general are represented in FIGS. 9A through 9F. FIGS. 9A, 9B and 9C are an example of deterministic vectors. This set of vectors is generated when the program characteristics are well defined, as
5 can occur when the source related text 136 or the EPG 140 contains specific fields identifying the category and the sub-category. A program rating can also provided by the EPG 140.

In the case that these characteristics are not specified, a statistical set of vectors is generated from the process
10 described in accordance with FIG. 8. FIGS. 9D-9F illustrate the probability that a program being watched is from the given category and program classification, respectively.

FIG. 10A illustrates sets of logical heuristics rules which form part of the heuristic rules 160. In a preferred
15 embodiment, logical heuristic rules are obtained from sociological or psychological studies. Two types of rules are illustrated in FIG. 10A. The first type links an individual's viewing characteristics to demographic characteristics such as gender, age, and income level. A channel changing rate rule
20 1030 attempts to determine gender based on channel change rate. An income related channel change rate rule 1010 attempts to link channel change rates to income brackets. A second type of rules links particular programs to particular audience, as illustrated by a gender determining rule 1050 which links the
25 program category/sub-category with a gender. The result of the application of the logical heuristic rules illustrated in FIG. 10A are probabilistic determinations of factors including gender, age, and income level. Although a specific set of logical heuristic rules has been used as an example, a wide
30 number of types of logical heuristic rules can be used to realize the present invention. In addition, these rules can be changed based on learning within the system or based on external studies which provide more accurate rules.

FIG. 10B illustrates a set of the heuristic rules

rule. This probability is the sum of the conditional probabilities that the demographic group likes the program, conditional to the category 144 weighted by the probability that the program is from that category 144. In a preferred embodiment, the program target analysis can calculate the program demographic vectors by the application of logical heuristic rules 160, as illustrated in FIG. 10A, and by the application of heuristic rules expressed as conditional probabilities as shown in FIG. 10B. Logical heuristic rules can be applied using logical programming and fuzzy logic using techniques well understood by those skilled in the art, and are discussed in the text by S. V. Kartalopoulos entitled "Understanding Neural Networks and Fuzzy Logic."

Conditional probabilities can be applied by simple mathematical operations multiplying program context vectors by matrices of conditional probabilities. By performing this process over all the demographic groups, the target analysis process 1100 can measure how likely a program is to be of interest to each demographic group. Those probabilities values form the demographic vector 170 represented in FIG.12.

As an example, the heuristic rules expressed as conditional probabilities shown in FIG. 10B are used as part of a matrix multiplication in which the program characteristics vector 150 of dimension N, such as those shown in FIGS. 9A-9F is multiplied by an N x M matrix of heuristic rules expressed as conditional probabilities, such as that shown in FIG. 10B. The resulting vector of dimension M is a weighted average of the conditional probabilities for each category and represents the household demographic characteristics 190. Similar processing can be performed at the sub-category and content levels.

FIG. 13 illustrates an entity-relationship diagram for the generation of household session demographic data 1310 and a household session interest profile 1320. In a preferred

actual subscriber selection data 199, and the heuristic rules 160 are applied. In the case of logical heuristic rules as shown in FIG. 10A, logical programming can be applied to make determinations regarding the household session demographic data 1310 and the household session interest profile 1320. In the case of heuristic rules in the form of conditional probabilities such as those illustrated in FIG. 10B, a dot product of the time averaged values of the program characteristics vectors 150 can be taken with the appropriate matrix of heuristic rules 160 to generate both the household session demographic data 1310 and the household session interest profile 1320.

Volume control measurements which form part of the subscriber selection data 110A and 110B can also be applied in the session characterization process 1300 to form a household session interest profile 1320. This can be accomplished by using normalized volume measurements in a weighted average manner similar to how time duration is used. Thus, muting a show results in a zero value for volume, and the program characteristics vector 150 for this show will not be averaged into the household session interest profile 1320.

FIG. 14 illustrates an entity-relationship diagram for the generation of average household demographic characteristics 190 and session household demographic characteristics 190. A household demographic characterization process 1400 generates the household demographic characteristics 190 represented in table format in FIG. 15. The household demographic characterization process 1400 uses the household viewing habits 195 in combination with the heuristic rules 160 to determine demographic data. For example, a household with a number of minutes watched of zero during the day may indicate a household with two working adults. Both logical heuristic rules as well as rules based on conditional probabilities can be applied to the household viewing habits 195 to obtain the household

the household interest profile 180 which takes into account both the viewing preferences of the household as well as assumptions about households/subscribers with those viewing habits and program preferences.

- 5 FIG. 17 illustrates an exemplary household interest profile 180 that includes a programming types row 1709, a products types row 1707, a household interests column 1701, an average value column 1703, and a session value column 1705.

10 The product types row 1707 gives an indication as to what type of advertisement the household would be interested in watching, thus indicating what types of products could potentially be advertised with a high probability of the advertisement being watched in its entirety. The programming types row 1709 suggests what kind of programming the household
15 is likely to be interested in watching. The household interests column 1701 specifies the types of programming and products which are statistically characterized for that household.

20 As an example of the industrial applicability of the invention, a household will perform its normal viewing routine without being requested to answer specific questions regarding likes and dislikes. Children may watch television in the morning in the household, and may change channels during commercials, or not at all. The television may remain off
25 during the working day, while the children are at school and day care, and be turned on again in the evening, at which time the parents may "surf" channels, mute the television during commercials, and ultimately watch one or two hours of broadcast programming. The present invention provides the ability to
30 characterize the household, and may make the determination that there are children and adults in the household, with program and product interests indicated in the household interest profile 180 corresponding to a family of that composition. A household with two retired adults will have a completely

is displayed on PC 1820 or television 1810 according to which device made the request.

According to one embodiment, the server side maintains the advertising selection data 110A and subscriber selection data 110B which it is able to compile based on its operation as a proxy for the client side. Retrieval of source related information 130, the program target analysis process 1100, the program characterization process 800, the session characterization process 1300, the household demographic characterization process 1400, and the household interest profile generation process 1600 can be performed by server 1840.

Referring to FIG. 19 an advertisement monitoring table is illustrated, in which an advertisement ID (AD ID) column 1915 contains a numerical ID for an advertisement which was transmitted with the advertisement in the form of a Program ID, http address, or other identifier which is uniquely associated with the advertisement. A product column 1921 contains a product description which indicates the type of product that was advertised. A brand column 1927 indicates the brand name of the product or can alternatively list a generic name for that product. A percent watched column 1933 indicates the percentage of the advertisement the subscriber viewed. In an alternate embodiment, a letter rating or other type of rating is used to indicate the probability that the advertisement was watched. A volume column 1937 indicates the volume level at which the advertisement was watched.

As an example of the industrial applicability of the invention, a manufacturer may develop an advertising strategy which includes the insertion of advertisements during popular evening programs. The costs for such ad insertions can be extremely high. In order to insure the cost effectiveness of this advertising strategy, the manufacturer has the advertisements placed during less watched but similar programs

Claims

What is claimed is:

1. A method for generating a subscriber profile based on advertisements watched by the subscriber, the method
5 comprising:

monitoring subscriber advertising viewing activities;

collecting advertisement selection data from the monitored
subscriber advertising viewing activities; and

generating a subscriber profile based on the advertisement
10 selection data.
2. The method of claim 1, further comprising retrieving
advertisement related information for advertisements identified
in the advertisement selection data wherein the advertisement
15 related information contains descriptive fields related to the
advertisements.
3. The method of claim 2, wherein said retrieving
includes context mining text associated with the
20 advertisements.
4. The method of claim 3, wherein the text is derived
from closed-captioning data associated with the advertisement.

collecting program selection data for the monitored subscriber program viewing activities;

combining the program selection data with the advertisement selection data;

5 evaluating the combined program selection data and the advertisement selection data to filter out irrelevant data and generate a record of actual subscriber selection data; and

processing the actual subscriber selection data to create the subscriber profile.

10

12. The method of claim 11, wherein said monitoring subscriber program viewing activities includes monitoring viewing time durations for selected source material.

15 13. The method of claim 11, wherein said evaluating the combined program selection data includes evaluating channel change commands and associated viewing times.

14. The method of claim 13, wherein said evaluating the
20 combined program selection data includes filtering out any channel change commands if the associated viewing times are below a pre-determined threshold.

20. The method of claim 11, wherein the subscriber profile is a product preference profile for the subscriber, the product preference profile indicating the type of products of interest to the subscriber.

5

21. The method of claim 11, wherein the subscriber profile is an advertising preference profile for the subscriber, the advertising preference profile indicating the type of advertising of interest to the subscriber.

10

22. A data processing system for generating a subscriber profile based on advertisements watched by the subscriber, the system comprising:

a storage medium;

15 means for monitoring subscriber advertising activities;

means for collecting advertisement selection data for the monitored subscriber advertising activities;

means for retrieving advertisement related information for ads in the advertisement selection data, wherein the

20 advertisement related information contains descriptive fields corresponding to the advertisements;

means for determining the extent to which the advertisements are viewed by the subscriber; and

27. The system of claim 26, wherein the text derived from closed-captioning data associated with the advertisement includes a product name field.

5 28. The system of claim 26, wherein the text derived from closed-captioning data associated with the advertisement includes a product brand field.

29. The system of claim 22, further comprising:
10 means for monitoring subscriber program viewing activities;
 means for collecting program selection data for the monitored subscriber program viewing activities; and
 means for combining the advertisement selection data and
15 the program selection data.

30. The system of claim 29, further comprising:
 means for filtering out irrelevant data and generating a record of actual subscriber selection data; and
20 means for processing the actual subscriber selection data to create a subscriber profile.

a storage medium;

means at the server side for monitoring subscriber activity including receiving subscriber channel change requests and storing subscriber channel change requests;

5 means at the server side for retrieving advertisement related information wherein the advertisement related information contains descriptive fields corresponding to an advertisement;

means at the server side for determining the extent to.
10 which an advertisement is viewed by the subscriber;

means at the server side for storing the descriptive fields and the determination of the extent to which the advertisement is viewed by the subscriber; and

means at the server side for processing the descriptive
15 fields and the extent to which the advertisement is viewed by the subscriber information to generate subscriber profile.

35. The system of claim 34, wherein the means for
retrieving advertisement related information further comprises
20 means for context mining of textual information associated with the selected source material.

means for processing the actual subscriber selection data to create a subscriber profile.

41. The system described in claim 34, wherein the
5 subscriber profile is a program preference profile for the subscriber, the program preference profile indicating the type of programming of interest to the subscriber.

42. The system of claim 34, wherein the subscriber
10 profile is a product preference profile for the subscriber, the product preference profile indicating the type of products of interest to the subscriber.

43. The system of claim 34, wherein the subscriber
15 profile is an advertising preference profile for the subscriber, the advertising preference profile indicating the type of advertising of interest to the subscriber.

2/27

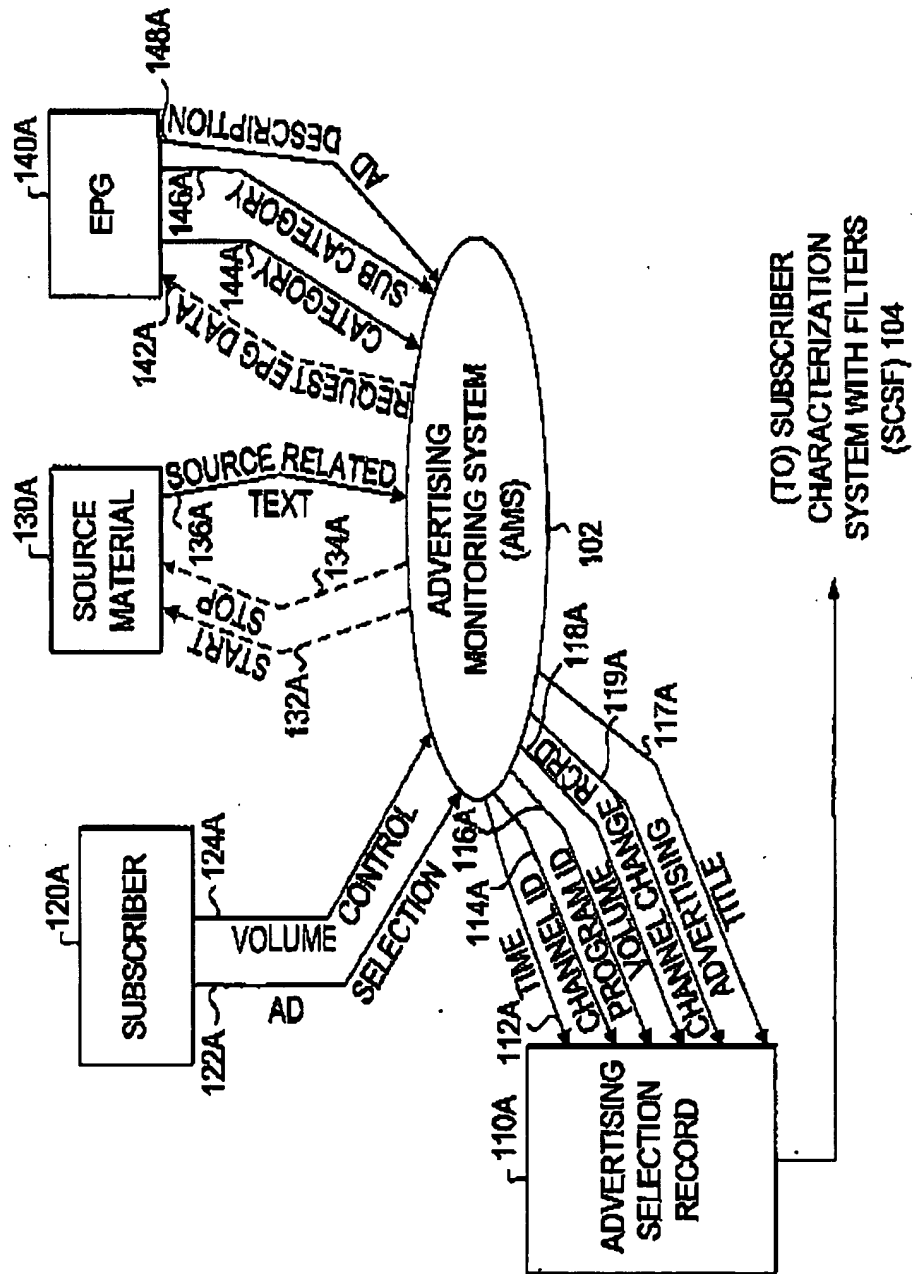


FIG. 1B

4/27

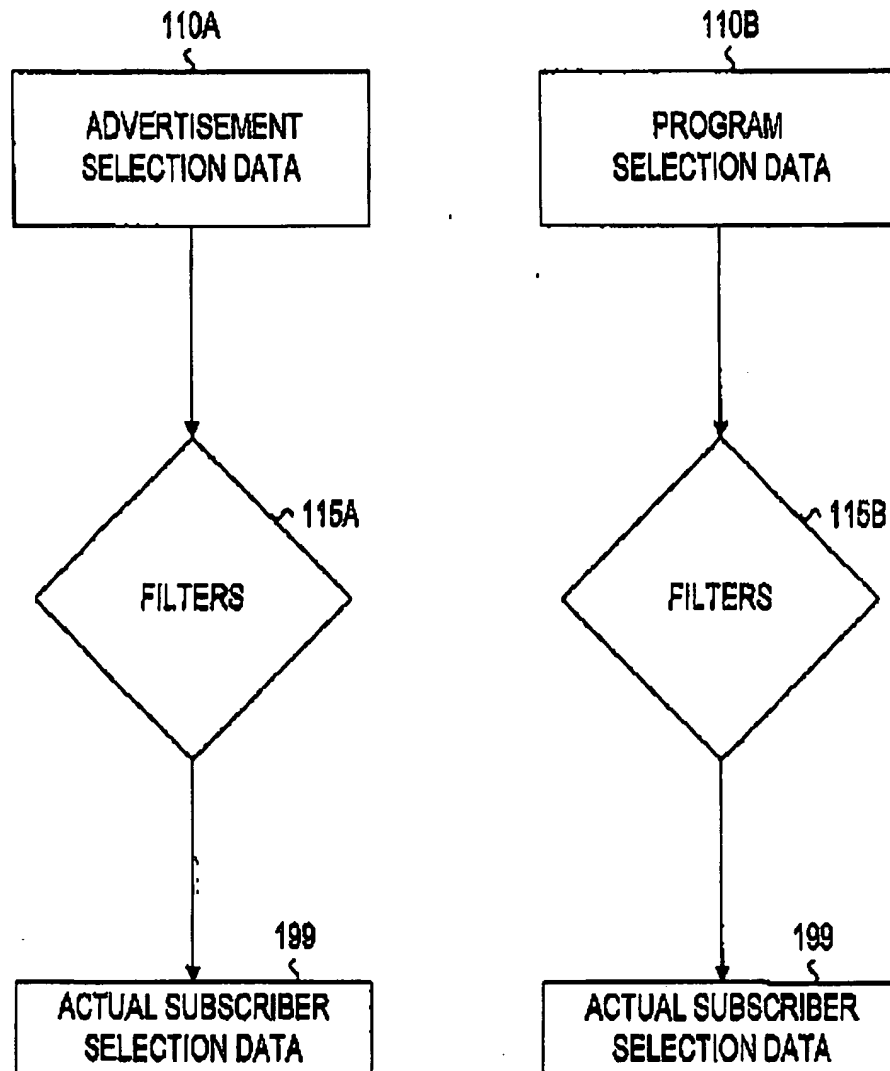


FIG. 1D

6/27

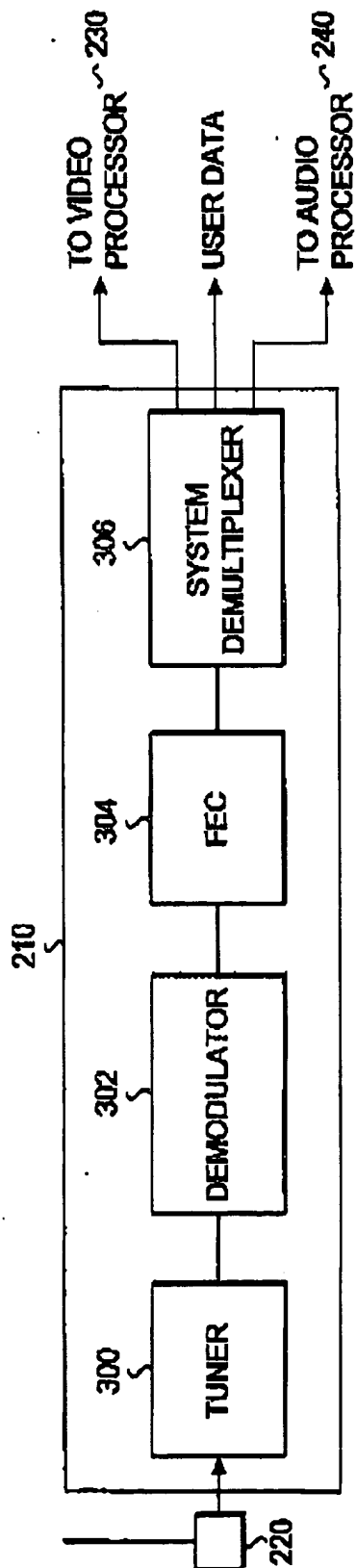


FIG. 3

8/27

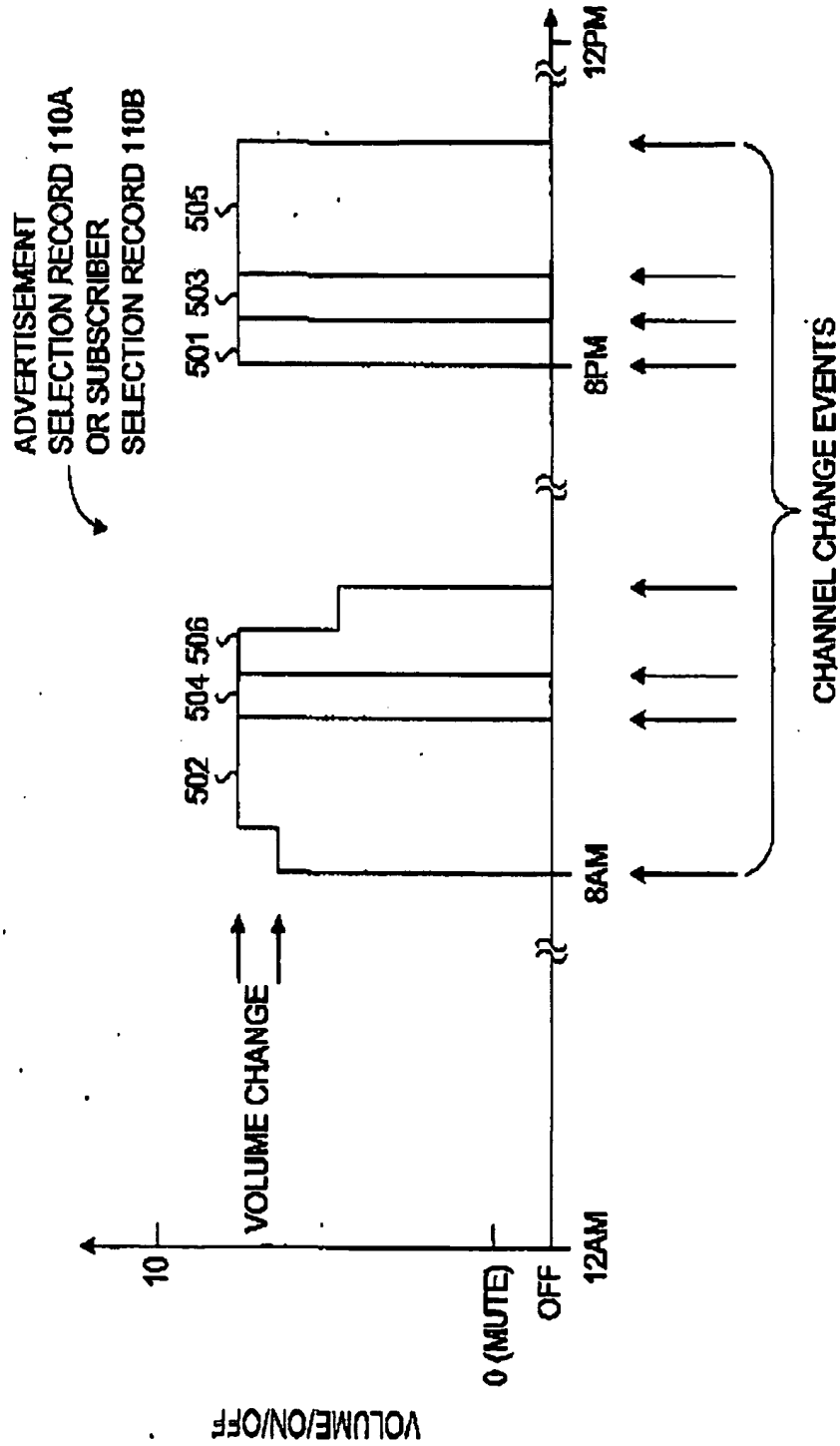


FIG. 5

10/27

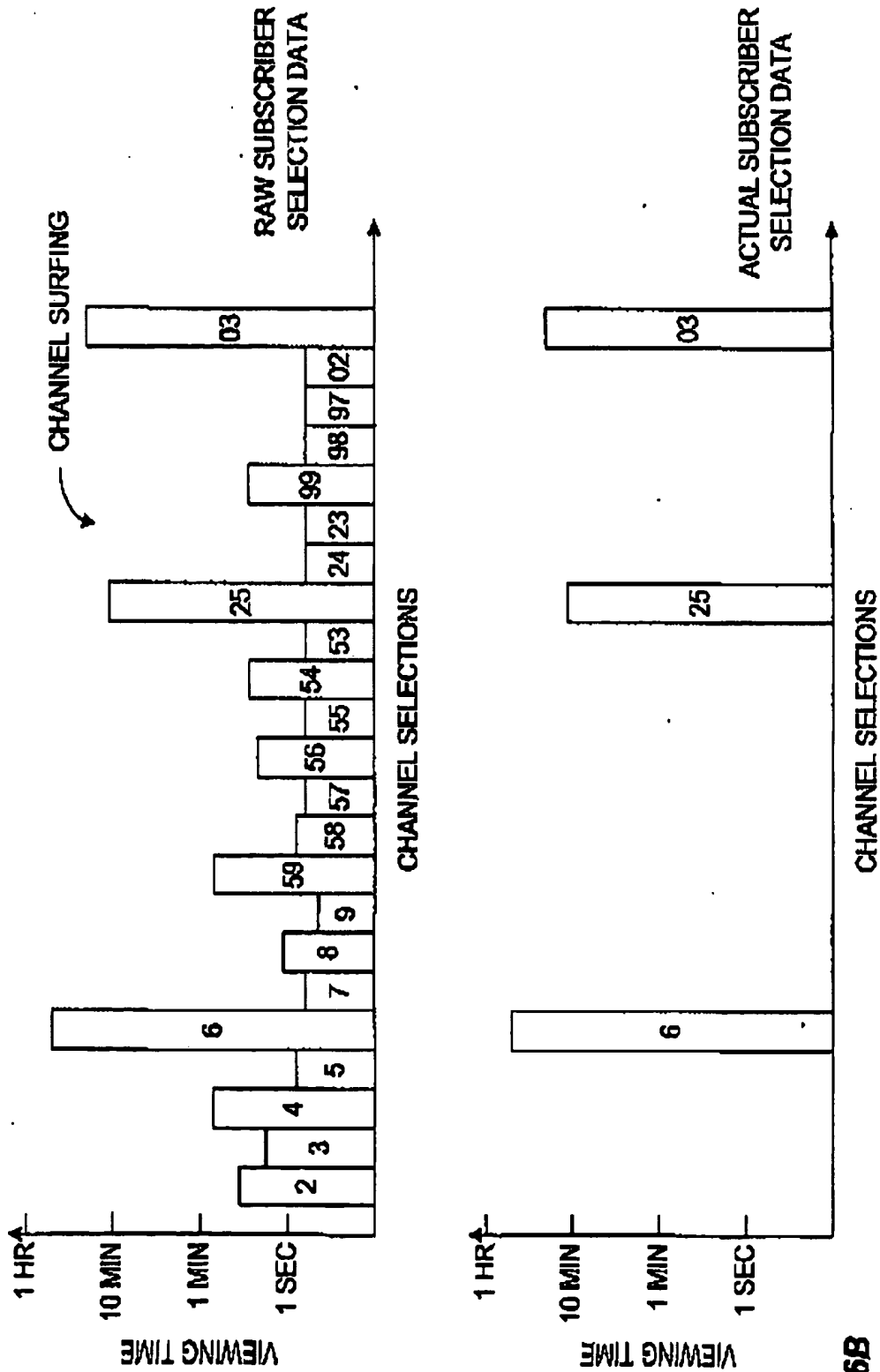


FIG. 6B

12/27

TIME OF DAY	MINUTES WATCHED	CHANNEL CHANGES	AVERAGE VOLUME
MORNING (6AM-9AM)	61	2	5/10
MID-DAY (9AM-3PM)	0	0	-
AFTERNOON (3PM-6PM)	0	0	-
NIGHT (6PM-10PM)	122	4	6/10
LATE NIGHT (12AM-6AM)	0	0	-
TOTAL	183	6	5.7/10

FIG. 7

14/27

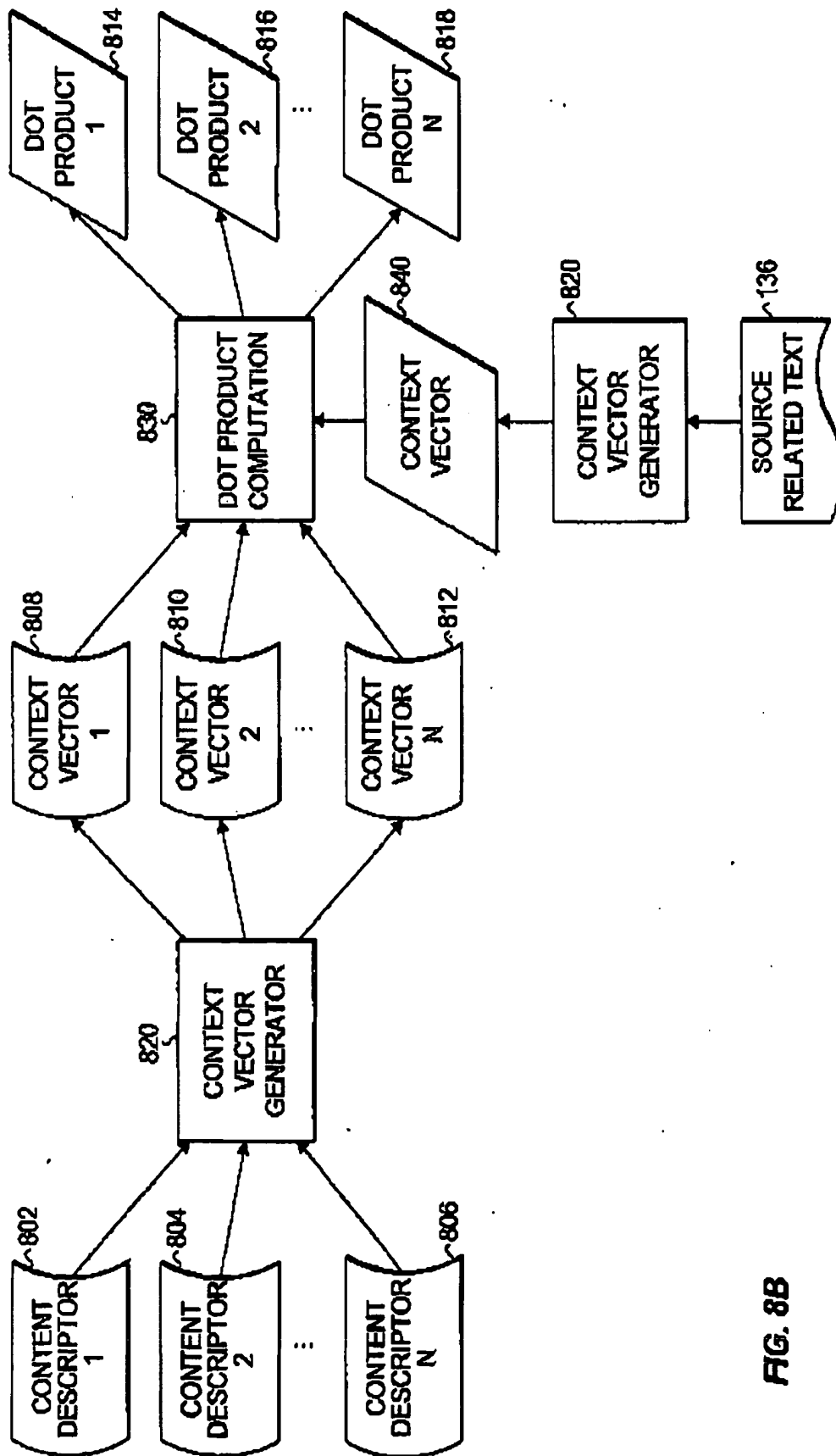


FIG. 8B

16/27

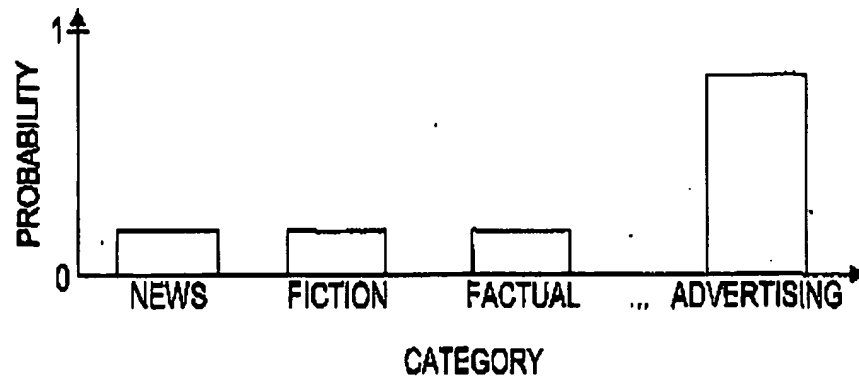


FIG. 9D

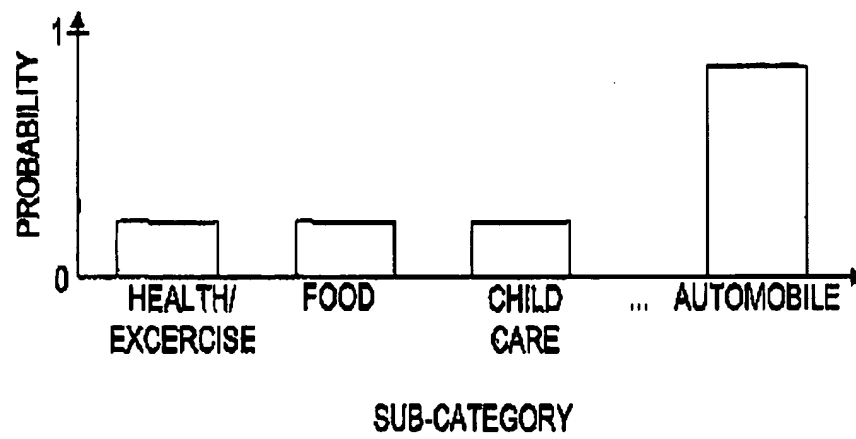


FIG. 9E

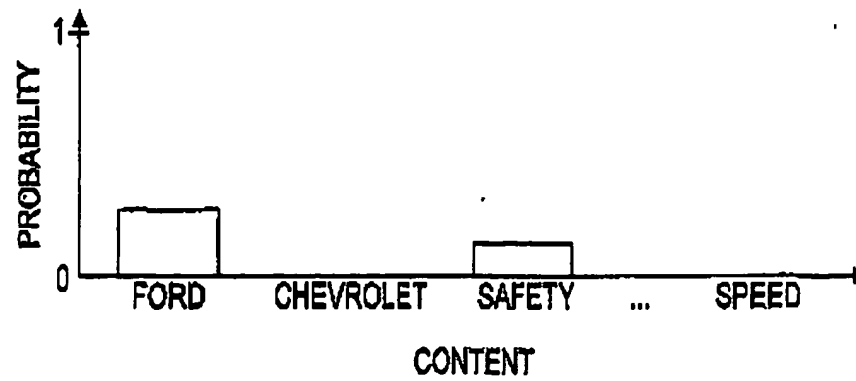


FIG. 9F

18/27

DEMOGRAPHIC GROUPS												
AGE			INCOME				SIZE			GENDER		
0-10 10-18 ... >70			0-20K 20-50K ... 50-100K				1 2 ... >5			M F		
CATEGORIES	NEWS	0.1	0.1	0.4	0.2	0.3	0.4	0.5	0.3	0.1	0.3	0.7
	FICTION	0.5	0.3	0.2	0.4	0.2	0.3	0.3	0.2	0.1	0.8	0.2
	FACTUAL	0.2	0.2	0.3	0.1	0.4	0.2	0.2	0.2	0.4	0.4	0.6
	ADVERTISING	0.1	0.3	0.5	0.3	0.2	0.1	0.2	0.1	0.3	0.5	0.5

FIG. 10B

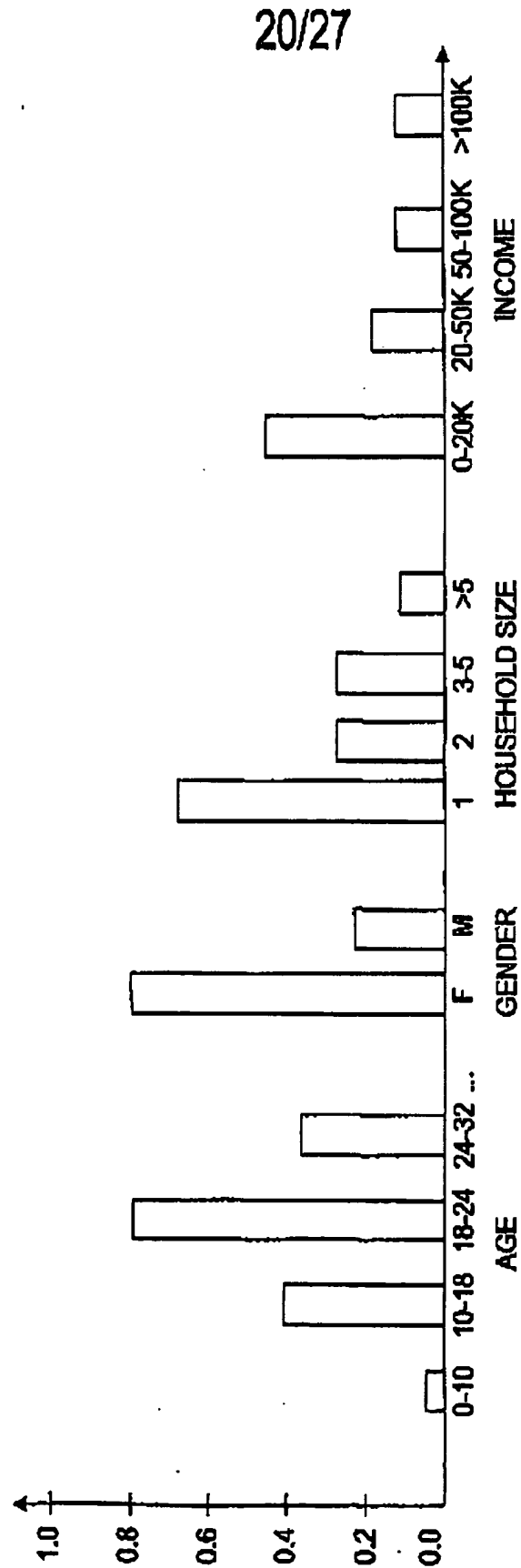


FIG. 12

22/27

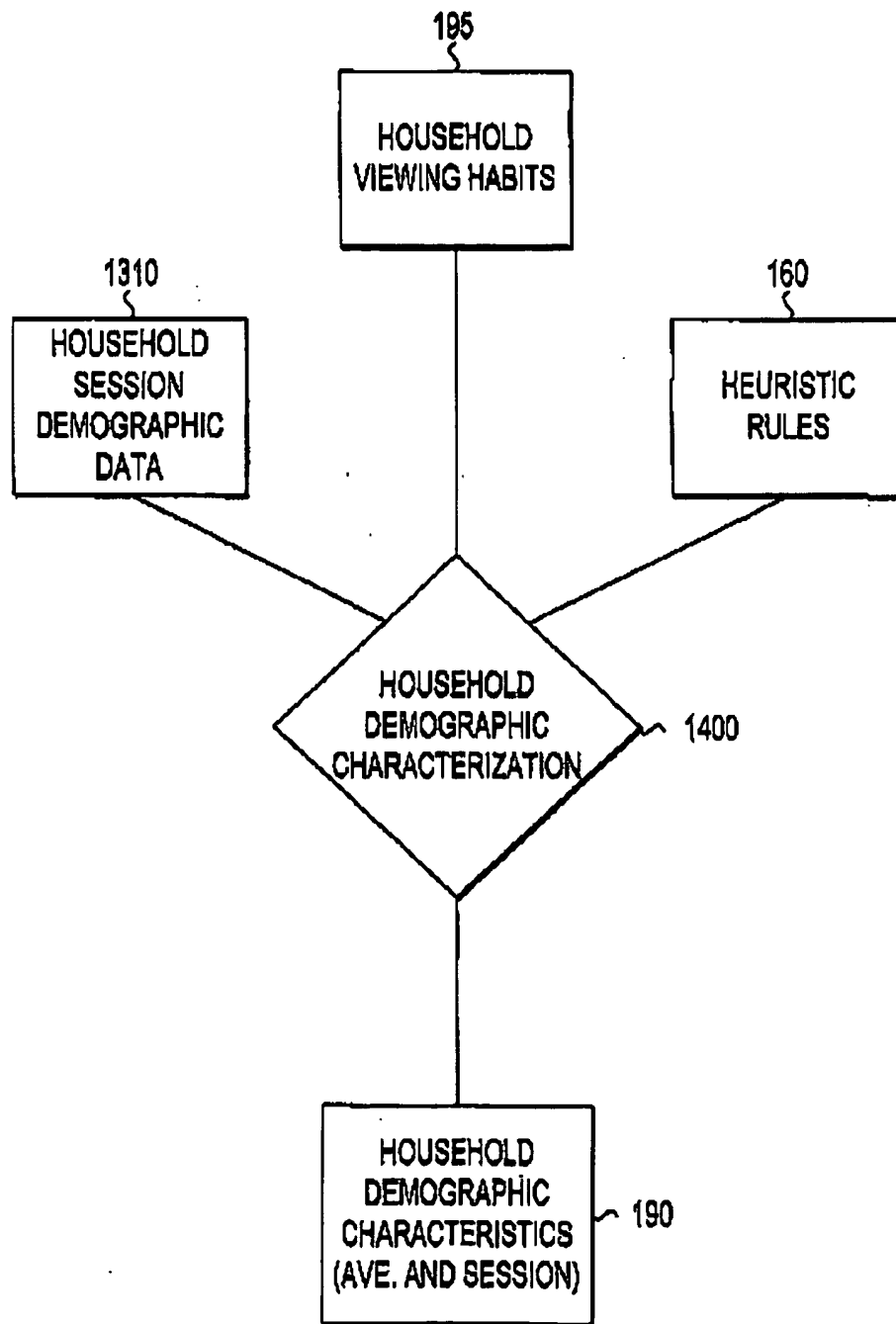


FIG. 14

24/27

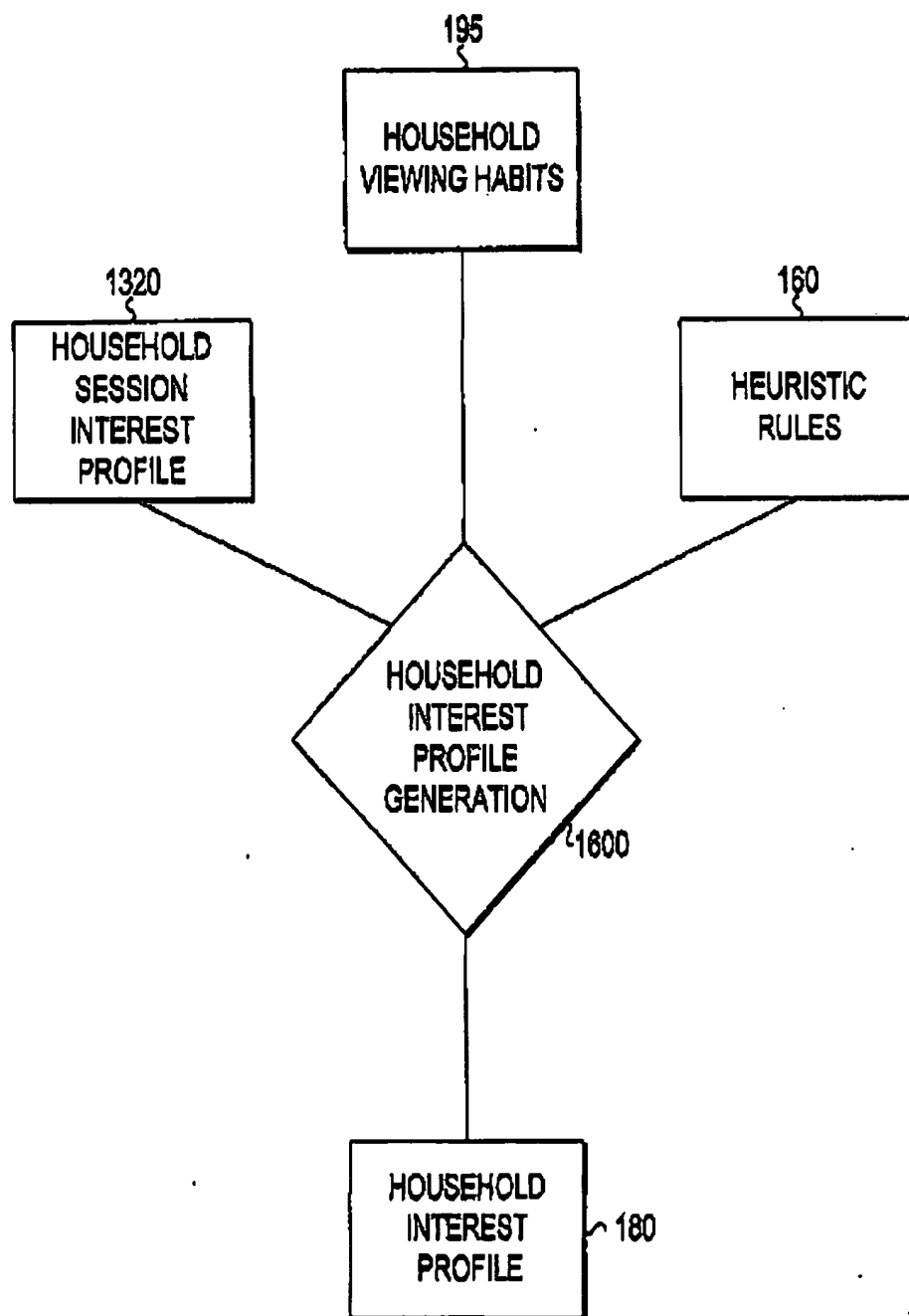


FIG. 16

26/27

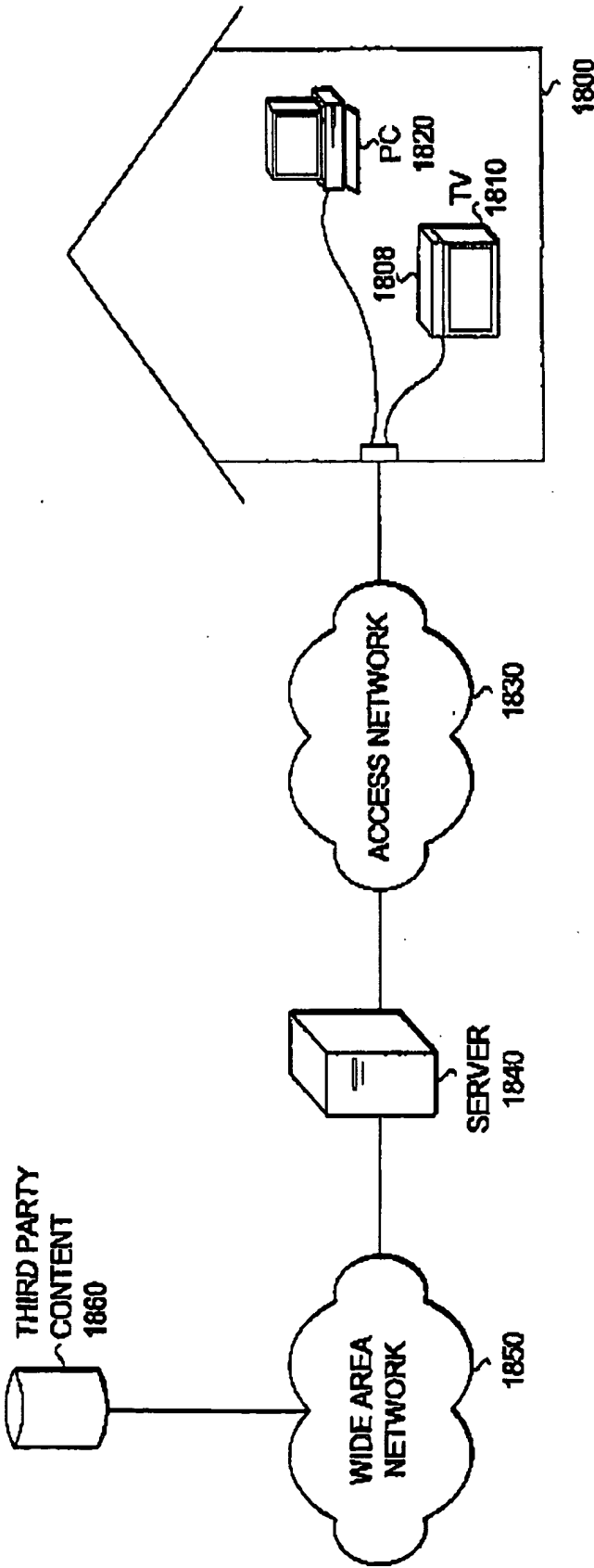


FIG. 18

INTERNATIONAL SEARCH REPORT

International application No.

PCT/U801/06459

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : H04H 9/00; H04N 7/025, 7/10, 7/16

US CL : 725/14, 34

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 725/9, 14, 15, 16, 17, 18, 19, 20, 22, 32, 34, 36

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,758,257 A (HERZ et al) 26 May 1998, see whole document.	1-43
Y	US 5,774,170 A (HITE et al) 30 June 1998, see whole document.	1-43



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	* ¹	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention.
A document defining the general state of the art which is not considered to be of particular relevance	* ²	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone.
B earlier document published on or after the international filing date	* ³	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
L document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	* ⁴	document member of the same patent family.
D document referring to an oral disclosure, use, exhibition or other means		
F document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

01 JUNE 2001

Date of mailing of the international search report

17 JUL 2001

Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-9230

Authorized officer

JOHN W. MILLER

Telephone No. (703) 305-4795

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.